DOCUMENT RESUME

ED 443 887 TM 031 532

AUTHOR Tomoff, Joan; Thompson, Marilyn; Behrens, John
TITLE Measuring NCTM-Recommended Practices and Student Achievement with TIMSS.
PUB DATE 2000-04-26
NOTE 15p.; Paper presented at the Annual Meeting of the American Educational Research Association (New Orleans, LA, April 24-28, 2000).
PUB TYPE Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Educational Practices; Factor Analysis; International Studies; *Junior High School Students; Junior High Schools; *Mathematics Achievement; Mathematics Education; Mathematics Tests; *Performance Factors; Secondary School Teachers; *Teacher Effectiveness
IDENTIFIERS *National Council of Teachers of Mathematics; *Third International Mathematics and Science Study

ABSTRACT

Data from Population 2 of the Third International Mathematics and Science Study (TIMSS) were used to explore relations between teacher classroom practices and student achievement. The sample consisted of 3,400 students and their 319 teachers. Principal components analysis focused classroom practices into four factors: (1) project creation; (2) group work; (3) drill and practice; and (4) textbook use. Hierarchical linear models were estimated to evaluate relationships between these factors and student mathematics test scores. Frequencies of project creation and of group work, two practices recommended by the National Council of Teachers of Mathematics (NCTM), were unrelated to scores on both a problem-solving subtest and the entire mathematics achievement test. Among the more traditional practices, the drill and practice factor had a negative relationship with overall test score, while working from a textbook was positively related to overall achievement. A student-level variable, highest parental educational attainment, was the strongest predictor of achievement scores. (Contains 5 tables and 16 references.) (Author/SLD)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 443 887

Measuring NCTM-Recommended Practices and Student Achievement with TIMSS

Joan Tomoff, Marilyn Thompson, and John Behrens

Arizona State University

April 26, 2000

Presented at AERA, New Orleans, Louisiana

Abstract

Data from Population 2 of the Third International Mathematics and Science Study (TIMSS) were used to explore relations between teacher classroom practices and student achievement. Principal components analysis focused classroom practices into four factors: project creation, group work, drill and practice, and textbook use. Hierarchical linear models were estimated to evaluate relationships between these factors and student mathematics test scores.

Frequencies of project creation and of group work, two practices recommended by the National Council of Teachers of Mathematics (NCTM), were unrelated to scores on both a problem-solving subtest and the entire mathematics achievement test. Among the more traditional practices, the drill and practice factor had a negative relationship with overall test score, while working from a textbook was positively related to overall achievement. A student-level variable, highest parental educational attainment, was the strongest predictor of achievement scores.

TM031532

BEST COPY AVAILABLE

(PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY)

J. Tomoff

(TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC))

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Measuring NCTM-Recommended Practices and Student Achievement with TIMSS

The implicit purpose of school mathematics has been expanding. Traditionally, intermediate and advanced school mathematics classes have been gatekeepers to higher education and a limited number of exclusive jobs, so the mathematical deficits of many students went unchallenged. With the increasing importance of technology, however, more good jobs require a working knowledge of mathematics, and more students need to learn problem solving and other quantitative workplace skills while in school. Despite a tradition of local control, many U.S. citizens argue for a common national set of performance standards, which may call for well-specified teaching standards and assessments. Our goal in this study was to explore the relationship between mathematics teaching standards and student performance, especially in problem solving, on a low-stakes standardized test.

To evaluate mathematics teaching and student achievement, we needed standards and a standardized test. An optional set of mathematics education standards, Curriculum and Evaluation Standards for School Mathematics (NCTM, 1989) has been available for a decade. Current research supports the standards' theoretical base, but most studies of their implementation and effects have been conducted in classrooms on a small scale, perhaps limiting their generalizability. If NCTM-recommended practices have the potential to revitalize mathematics instruction across the nation, they must be supported and studied on a large scale. The Third International Mathematics and Science Survey (TIMSS) provides a large sample of pertinent student and teacher information and is capable of supporting a multilevel analysis relating teacher practices to student achievement. Of particular interest in this study are student background data and standardized mathematics achievement test scores, accompanied by teachers' reports of their instructional practices.

This study addresses two hypotheses. First, self-reported teacher classroom practices reflect a small number of well-defined factors. Second, these classroom practice factors can be related to student responses on achievement test items. If we can show that students of teachers using NCTM-recommended practices perform at least as well as students of more traditional teachers on most aspects of a standardized test, but perhaps better on problem solving, then the need for mathematics classroom reform will be supported.

Method

Sample

The middle-grade group of the TIMSS data set, Population 2, was used in this study. From each school included in the study, students were sampled by intact mathematics classrooms. By using two 8th- grade and one 7th-grade classroom from each school, student achievement and background data could be matched with the instructional practices reported by their teachers. After eliminating students from the sample due to missing scores on variables of interest at the student level or due to missing responses from their teachers, the sample consisted of 3400 students and their 319 teachers.

Analysis of Teacher Practices

The teacher questionnaire contained hundreds of variables addressing the teacher's background, educational beliefs, and instructional practices. From these, 40 were selected for their potential to distinguish traditional practices from NCTM-compatible practices. A series of exploratory principal components analyses were conducted to identify teacher practice factors. Tomoff (1999) presents the rationale and detailed procedures for these analyses.

Analysis of Student Achievement as Related to Teacher Practices

The relationships between student achievement and teacher factors were explored via a series of hierarchical linear models for predicting student achievement in mathematics. Two-level models were specified: students (level 1) were nested in classrooms (level 2). The models were estimated using HLM Version 4.04 (Bryk, Raudenbush, & Congdon, 1996). Residual files were checked, and results supported the assumption that random effects were normally distributed.

In a preliminary study, student scores on problem-solving items were modeled separately from student scores on multiple choice items concerning knowledge and computation. Effects of parental education were stronger for the problem-solving items, but traditional teacher classroom practice factors had stronger and more significant effects on international scores. For the final analysis, therefore, we decided to retain the reliability and validity of the test as developed and to use overall test score (INTSCR) as the indicator for mathematics achievement outcome. The unconditional model, containing no explanatory variables, was first estimated as a baseline for comparison with models containing one or more explanatory variables. Highest parent educational attainment was added as a demographic control variable to predict student achievement. Models were then estimated that included separate sets of teacher practice conditions: NCTM-recommended practices and traditional practices. Finally, the complete model with all student and teacher conditions was evaluated.

Results

Defining Teacher Practices

Correlations among 27 of the 40 variables chosen from the teacher questionnaire were slight. However, principal components analysis of the remaining 13 variables with Promax rotation yielded four teacher practice factors: creating projects and reports, working in large and small groups, practicing algorithms, doing exercises from textbooks and worksheets. Table 1 shows that five items loaded on the projects factor, three each on the group work and practice factors, and two on the text-based teaching factor. Table 2 shows that four factors explain just over 60% of the variance of those thirteen items.

Table 3, the factor correlation matrix, shows that correlations between the two NCTM-compatible factors and between the two traditional practice factors were relatively small but statistically significant. The low correlation between the project and group work factors, which indicate practices compatible with the NCTM Standards, may suggest that teachers implement some but not all aspects of the standards. The algorithms and textbook factors, which indicate a more traditional and behavioristic teaching style, and were less highly correlated than expected.

Relating Teacher Practices to Student Achievement

Several hierarchical linear models were used to explore the relationship between teacher classroom practices and student achievement. The models are summarized in the following text, and the model statistics reported in Tables 4 and 5 facilitate comparisons among models.

The unconditional model at the student level shows individual international score as a function of the classroom mean international score (the intercept) plus student-level random error:

$$INTSCR_{ij} = MINTSCR_j + e_{ij} \quad \text{or} \quad Y_{ij} = \beta_{0j} + e_{ij}, \quad e_{ij} \sim N(0, \sigma^2) \qquad (1)$$

At the classroom level, mean international score is a function of the grand mean of classroom international scores plus classroom-level random error:

$$MINTSCR_j = GMINTSCR + \delta_{0j} \text{ or } \beta_{0j} = \gamma_{00} + \delta_{0j}, \quad \delta_{0j} \sim N(0, \tau_0) \qquad (2)$$

The unconditional model serves as a baseline for comparison with models containing explanatory variables.

The student model, with one predictor representing highest parental educational attainment (HPED), was compared to the unconditional model. Each student HPED score was centered about their respective class mean parental educational attainment ($MHPED_j$) to create the deviated predictor (DHPED). This centering approach makes the fixed effects more readily interpretable. With this parameterization, the intercept β_{0j} is the predicted classroom mean achievement score for students at their classroom mean on HPED. The classroom mean educational attainment (MHPED) is then added back into the model at the classroom level to make the fixed effects orthogonal across the student and classroom levels.

At the student level, $$Y_{ij} = \beta_{0j} + \beta_{1j} DHPED_{ij} + e_{ij}. \qquad (3)$$

At the classroom level, $$\beta_{0j} = \gamma_{00} + \gamma_{01} MHPED_j + \delta_{0j}. \qquad (4)$$

and $$\beta_{1j} = \gamma_{10} + \gamma_{11} MHPED_j + \delta_{1j}. \qquad (5)$$

As shown in Table 4, variance (σ^2) at the student level decreased from the unconditional model by only 2% [(4384.96 – 4283.34)/4384.96]. At the classroom level, however, the variance associated with the first intercept (τ_0) decreased by nearly 35% [(4344.78 –

2847.47)/4344.78]. This suggests parental education level (HPED) accounted for 35% of the between-classroom variance in mathematics achievement.

As shown in Table 5, the effect of the overall classroom mean score was significant and positive ($\gamma_{00} = 250.60$, $p < .01$), and average educational attainment of a classroom of students' parents was positively correlated with the outcomes of students within that classroom ($\gamma_{01} = 52.36$, $p < .01$). That is, students in classes with higher average scores and better-educated parents had significantly higher mathematics achievement, as measured by their international scores. Students in classes with lower scores and less educated parents had lower mathematics international scores. Average level of parental education in a classroom did not significantly affect the relationship between parental education and TIMSS international scores ($\gamma_{11} = 9.65$, $p = .22$).

In the full model with all four teacher classroom practice factors, neither set of NCTM-recommended practices significantly affected international scores. However, the two sets of traditional practices did have statistically significant effects. Therefore, a model was developed and analyzed with these two factors as classroom level predictors; the centered parental education variable was also maintained in this model.

At the student level, $$Y_{ij} = \beta_{0j} + \beta_{1j} DHPED_{ij} + e_{ij}. \quad (6)$$

At the classroom level, $$\beta_{0j} = \gamma_{00} + \gamma_{01} MHPED_j + \gamma_{04} ALGO_j + \gamma_{05} TEXT_j + \delta_{0j} \quad (7)$$

and $$\beta_{1j} = \gamma_{10} + \delta_{1j}. \quad (8)$$

The model that seemed to explain the most variance in mathematics achievement scores is illustrated in Equations 6, 7, and 8. The TIMMS international score was the student outcome, and highest parental educational attainment was the student level predictor. Classroom level predictors were mean parental education (MHPED), frequency of drill and

practice with algorithms and rules (ALGO) and percent of the class based on a textbook (TEXT). As shown in Table 4, the correlation between classroom average score and parental education level decreased slightly from that in the student model, from .35 to .31 or about 11%. Apparently teacher practices accounted for a small amount of score variance that had been attributed to parental education in the student model.

Final estimations of fixed effects for this model containing traditional practice predictors were compared to those of two other models and are shown in Table 5. As with the model containing only the parental education predictor, students in classes with higher average scores and better-educated parents had significantly higher mathematics achievement ($\gamma_{00} = 269.16$, $p < .01$ and $\gamma_{01} = 48.26$, $p < .01$, respectively). Frequency of drill and practice had a slight negative effect ($\gamma_{04} = -10.51$, $p < .01$), and percent of class based on a textbook had a slight positive effect on classroom mean scores ($\gamma_{05} = 6.77$, $p = .05$).

Discussion

Student outcomes were strongly related to parental educational attainment (HPED). In the preliminary study, parental education was correlated more highly with classroom average free response score (.51) than with classroom average international score (.35). That is, students with better-educated parents had higher mathematics test scores, and the effect was stronger for problem solving than for overall achievement. In addition, for the problem solving outcome, average level of parental education in a classroom enhanced the positive relationship between parental education and achievement, whereas for the international score outcome, this effect was absent. These results confirm the generally strong relation between socioeconomic indicators and test scores found by other researchers (e.g. Dossey, Mullis, Lindquist & Chambers, 1984; and Jones, 1987, cited in Secada, 1992).

Practice with rules and algorithms (ALGO) showed a negative relationship with test scores. Students who practiced the most computation were the least skillful. Most likely, teachers devote significant time to drill and practice when students have deficient knowledge of rules and facility with math facts. When students understand rules and are proficient with computation, teachers can profitably spend more time with problem solving and other useful mathematics not measured by standardized tests.

Percentage of the course based on the textbook (TEXT) showed a positive relationship with test scores. There are several possible explanations. Perhaps teachers who emphasize the role of the text spend more of each class period actively teaching. Other researchers have found direct correlations between student achievement and quantity of time teachers spend engaging students with subject matter (Wittrock, 1986). Perhaps students working from a text spend more time focused on learning than do students working with a variety of materials. Dependence on a textbook requires fewer class management skills than activity-based teaching with manipulatives and other materials. Teaching from a text also requires less teacher knowledge than teaching well using alternative methods. Finally, curriculum taught from a textbook may be more likely to be aligned with a standardized test, causing students with more book exposure to score higher. Perhaps more specialized curricula based on manipulatives and other non-text materials require special assessments.

While it is disheartening to find that teacher practices or instructional quality have minimal effects on TIMSS performance, others have found similar results when trying to study relationships between instructional practices and standardized test scores. For example, Young, Reynolds, and Walberg (1996) found that science achievement, as measured by test items on the National Assessment of Educational Progress (NAEP, 1986, cited in Young et

al) had a stronger relationship with home environment than with quality of instruction. Lee and Croninger (1994), measuring variables at the school rather than the classroom level, found that "authentic instruction in English classes" had no independent influence on mean achievement. This may partially be explained by the tendency for overall achievement measures to reflect cumulative educational attainment rather than gains resulting from practices of a particular teacher.

The most important limitation of the background data, students' parental education levels and teachers' classroom practices, is that they were self-reported. Although some researchers who have compared teachers' self-reports with classroom observations say such reports of classroom practices are reliable (Burstein et al., 1995), others with similar experience disagree (J. Knaupp, personal communication, June 16, 1999). Mayer (1999) found that a self-reported composite of classroom practices was both reliable and valid although individual indicators were less trustworthy. Because the classroom practice variables in this study were composites, they were assumed to be reliable. However, no available questionnaire can measure the quality of implementation of teaching practices, and this probably affects student learning more than any single practice factor.

An important implication is that standardized tests do not measure teaching. Yet policymakers, the public, and many educators believe test scores can be used to gauge the quality of teaching in a classroom, school, or district. Despite a movement toward more "authentic" assessment methods, standardized test scores remain frequently used and widely publicized tools measuring students' learning. The TIMSS took steps to include more authentic measures when they added free-response problem-solving items to a largely

multiple choice test. If there had been more such questions, and if all students had received them, our results might have been more definitive.

Can an important and complicated educational outcome like problem-solving ability be measured by a standardized test? Some testing experts (e.g. Haladyna, 1994) believe that properly developed multiple choice items can accurately measure complex cognitive behavior. If true problems can be placed in a multiple choice format, problem-solving assessments can be machine scored. However, careful attention must be given to the skills evaluated by the items. Some have argued that test construction frequently depends more on error structure than item content; that is, more emphasis is placed on statistical performance of items than on the concepts the items are testing (Haertel & Wiley, 1993). The TIMSS added one small section of problem solving items. If instruction in problem solving is valued, and if we desire to assess the effects of such instruction with standardized tests, high-quality problem solving items must be developed, evaluated, and included in large-scale assessments.

References

Bryk, A. S. & Raudenbush, S. W. (1992). Hierarchical linear models: Applications and data analysis methods. Newbury Park, CA: Sage.

Bryk, A. S., Raudenbush, S. W., & Congdon, R. T. (1996). HLM: Hierarchical linear and nonlinear modeling with the HLM/2L and HLM/3L programs. Chicago: Scientific Software International.

Burstein, L., McDonnell, L. M., Van Winkle, J., Ormseth, T., Mirocha, J., & Guitton, G. (1995). Validating national curriculum indicators. Santa Monica, CA: RAND.

Haertel, E. H. & Wiley, D. E. 1993. Representations of ability structures: Implications for testing. In N. Fredericksen, R. J. Mislevy, & I. I. Bejar (Eds.), Test theory for a new generation of tests (pp. 359-384). Hillsdale, NJ: Lawrence Erlbaum Associates.

Haladyna, T. 1994. Developing and validating multiple choice test items. Hillsdale, NJ: Lawrence Erlbaum Associates.

Koretz, D., Stecher, B., Klein, S., & McCaffrey, D. (1994). The Vermont portfolio assessment program: Findings and implications. Educational Measurement: Issues and Practice, 13 (3), 5-16.

Lee, V. E. & Croninger, R. G. (1994). The relative importance of home and school in the development of literacy skills for middle-grade students. American Journal of Education 102 (3), 286-329.

Mayer, D. P. (1998). Do new teaching standards undermine performance on old tests? Educational Evaluation and Policy Analysis 20 (2), 53-73.

National Council of Teachers of Mathematics. (1989). Curriculum and evaluation standards for school mathematics. Reston, VA: National Council of Teachers of Mathematics.

Phelps, R. P. (1998). The demand for standardized student testing. Educational Measurement: Issues and Practice, 17 (3), 5-23.

Secada, W. G. (1992). Race, ethnicity, social class, language, and achievement in mathematics. In D. A. Grouws (Ed.), Handbook of research on mathematics teaching and learning (pp. 623-660). New York: Macmillan.

Smith, M. L. (1991). Put to the test: The effects of external testing on teachers. Educational Researcher, 20 (5), 8-11.

Tomoff, J. (1999). Can student achievement on standardized tests be related to NCTM-recommended teacher practices? Unpublished doctoral dissertation, Arizona State University.

Webb, N. L. (1992). Assessment of students' knowledge of mathematics: Steps toward a theory. In D. A. Grouws (Ed.), Handbook of research on mathematics teaching and learning (pp. 661-683). New York: Macmillan.

Wittrock, M. C. (1986). Students' thought processes. In M. C. Wittrock (Ed.), Handbook of research on teaching (3rd ed., pp. 297-314). New York: Macmillan.

Young, D. J., Reynolds, A. J., & Walberg, H. J. (1996). Science achievement and educational productivity: A hierarchical linear model. Journal of Educational Research, 89 (5), 272-278.

Table 1
Component Matrix of Teacher Classroom Practice Items

Item	Component			
	Projects	Groups	Practice	Text
Small investigations	.85			
Individual projects	.81	.12		
Small group projects	.85	.13		−.10
Find uses of content	.70		.10	
Prepare oral reports	.72		−.30	
Work together/teacher teaches	−.18	.71	.14	
Work together/students interact	.17	.80		−.12
Work in groups/assistance	.17	.68		
Practice to overcome difficulty		−.15	.73	
Practice computation		.15	.66	
Math: algorithms and rules			.67	.21
Textbook problems				.83
Textbook-based teaching	−.12			.84

Table 2
Rotation Sums of Squared Loadings of Teacher Classroom Practice Components

Component	Eigenvalue	% of Variance	Cumulative %
Projects	3.21	24.66	24.66
Groups	1.68	12.88	37.54
Practice algorithms	1.58	12.15	49.69
Text-based teaching	1.47	11.31	60.99

Table 3
Correlations Among Teacher Classroom Practice Factors

Factor	Projects	Groups	Algorithms	Text
Projects	1.00			
Groups	.13*	1.00		
Practice algorithms	−.09	.03	1.00	
Text-based teaching	−.10	−.08	.14**	1.00

*p = .02, ** p = .01

Table 4
Variances and Deviance Statistics. ($221.22 \leq Y_{ij} \leq 816.67$)

Model	σ^2	τ_0 (β_{0j})	τ_1 (HPED)
(1) Unconditional	4384.96	4344.78	
(2) Student	4283.34	2847.47	149.66
(4) Traditional	4281.05	2714.58	128.52

	τ as Correlations		Deviance	Estimated Parameters
(1) Unconditional			38883.40	2
(2) Student	τ_0	1.00	38737.42	4
	τ_1	.35		
(4) Traditional	τ_0	1.00	38720.20	4
	τ_1	.31		

Table 5
Estimations of Fixed Effects ($221.22 \leq Y_{ij} \leq 816.67$)

Model	Effect	Coefficient	t	p
(1) Unconditional:				
$Y_{ij} = \beta_{0j} + e_{ij}$	γ_{00}	488.88	124.39	.00
$\beta_{0j} = \gamma_{00} + \delta_{0j}$				
(2) Student:				
$Y_{ij} = \beta_{0j} + \beta_{1j}DHPED_{ij} + e_{ij}$	γ_{00}	250.60	12.20	.00
$\beta_{0j} = \gamma_{00} + \gamma_{01}MHPED_j + \delta_{0j}$	γ_{01}	52.36	11.77	.00
$\beta_{1j} = \gamma_{10} + \gamma_{11}MHPED_j + \delta_{1j}$	γ_{10}	9.65	1.22	.22
	γ_{11}	−1.31	−.75	.46
(4) Traditional:				
$Y_{ij} = \beta_{0j} + \beta_{1j}DHPED_{ij} + e_{ij}$	γ_{00}	269.16	12.58	.00
$\beta_{0j} = \gamma_{00} + \gamma_{01}MHPED_j +$	γ_{01}	48.26	10.39	.00
$\gamma_{04}ALGO_j + \gamma_{05}TEXT_j + \delta_{0j}$	γ_{04}	−10.51	−3.27	.00
$\beta_{1j} = \gamma_{10} + \delta_{1j}$	γ_{05}	6.77	1.99	.05
	γ_{10}	3.76	3.30	.00

n = 3400